04 - Properties of Multivariate Normal Distribution

Junvie Pailden

SIUE, F2017, Stat 589

September 05, 2017

Properties of Multivariate Normal Distribution

1. If $\mathbf{X} \sim N_p(\mu, \Sigma)$, then any linear combination of variables $\mathbf{a}'\mathbf{X} = a_1 X_1 + \cdots + a_p X_p \sim N(\mathbf{a}'\mu, \mathbf{a}'\Sigma\mathbf{a})$.

2. If $\mathbf{a}'\mathbf{X} \sim N(\mathbf{a}'\mu, \mathbf{a}'\Sigma\mathbf{a})$ for every $\mathbf{a}$, then $\mathbf{X}$ must be $N_p(\mu, \Sigma)$. (Alternative definition of MVN)

3. If $\mathbf{X} \sim N_p(\mu, \Sigma)$, the q linear combinations

$$\mathbf{A}_{(q\times p)}\mathbf{X}_{(p\times 1)} = \begin{bmatrix} a_{11}X_1 + \cdots + a_{1p}X_p \\ a_{21}X_1 + \cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \cdots + a_{qp}X_p \end{bmatrix} \sim N_q(\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}')$$

   All subsets of $\mathbf{X}$ are normally distributed.

Properties of Multivariate Normal Distribution

4. If $\mathbf{X}_1$ $(q_1 \times 1)$ and $\mathbf{X}_2$ $(q_2 \times 1)$ are independent, then $\Sigma_{12} = Cov(\mathbf{X}_1, \mathbf{X}_2) = 0$, a $q_1 \times q_2$ matrix of zeros.

5. If
$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N_{q_1+q_2}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right),$$
then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $\Sigma_{12} = 0$.

6. If X_1 and X_2 are independent and distributed as $N_{q_1}(\mu_1, \Sigma_{11})$ and $N_{q_2}(\mu_2, \Sigma_{22})$, then
$$\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N_{q_1+q_2}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \right)$$

Conditional Distribution

- Let $\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim N_{q_1+q_2}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$, and $|\Sigma_{22}| > 0$, where $\mathbf{X}_1$ $(q_1 \times 1)$ and $\mathbf{X}_2$ $(q_2 \times 1)$.

- Then the conditional distribution of $\mathbf{X}_1$, given that $\mathbf{X}_2 = \mathbf{x}_2$, is normal and has

$$
\begin{aligned}
E(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) &= \mu_{1|2} \\
&= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2) \\
Cov(\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2) &= \Sigma_{1|2} \\
&= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\
\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2 &\sim N_{q_1}(\mu_{1|2}, \Sigma_{1|2}) \ \ \text{(usual notation)}
\end{aligned}
$$

Multivariate Normal Conditional Distribution

1. All conditional distributions are (multivariate) normal.
2. The conditional mean is of the form

$$
\begin{aligned}
&\mu_1 + \beta_{1,q+1}(x_{q+1} - \mu_{q+1}) + \cdots + \beta_{1,p}(x_p - \mu_p) \\
&\vdots \\
&\mu_q + \beta_{q,q+1}(x_{q+1} - \mu_{q+1}) + \cdots + \beta_{q,p}(x_p - \mu_p)
\end{aligned}
$$

where the β's are defined by

$$
\Sigma_{12}\Sigma_{22}^{-1} = \begin{bmatrix} \beta_{1,q+1} & \beta_{1,q+2} & \cdots & \beta_{1,p} \\ \beta_{2,q+1} & \beta_{2,q+2} & & \beta_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{q,q+1} & \beta_{q,q+2} & \cdots & \beta_{p,q+1} \end{bmatrix}
$$

3. The conditional variance, $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, do not depend on the value(s) of the conditioning variable.

Conditional Density of a Bivariate Normal Distribution

1. If $f(x_1, x_2)$ is the bivariate normal density, show that $f(x_1|x_2)$ is
$$N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2),\ \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right).$$

   Example Bivariate Normal
$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_2\left(\begin{bmatrix} 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 1.2 & 1 \\ 1 & 1.2 \end{bmatrix}\right)$$

2. What is mean and variance of $X_1|X_2 = 4$?
3. What is the distribution of $X_1|X_2 = 4$?
4. Find $\Pr(X_1 \leq 3.5|X_2 = 4)$.

Spectral Decomposition (SD)

Let $\mathbf{A}$ be a $k \times k$ positive definite symmetric matrix, the SD of $\mathbf{A}$ is

$$\mathbf{A} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \cdots + \lambda_k \mathbf{e}_k \mathbf{e}_k' = \sum_{i=1}^{k} \lambda_i \mathbf{e}_i \mathbf{e}_i' = \mathbf{P} \boldsymbol{\Delta} \mathbf{P}'$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k > 0$ are eigenvalues of $\mathbf{A}$ and $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k$ are the associated normalized eigenvectors ($\mathbf{e}_i' \mathbf{e}_i = 1, \mathbf{e}_i' \mathbf{e}_j = 0$, for $i \neq j$).

$$\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k], \quad \mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \mathbf{I}$$

$$\boldsymbol{\Delta} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}$$

$$A^{-1} = \mathbf{P} \boldsymbol{\Delta}^{-1} \mathbf{P}' = \sum_{i=1}^{k} \frac{1}{\lambda_i} \mathbf{e}_i \mathbf{e}_i'$$

Square Root of a Positive Definite Matrix

$$A^{1/2} = \mathbf{P}\Delta^{1/2}\mathbf{P}' = \sum_{i=1}^{k} \sqrt{\lambda_i}\mathbf{e}_i\mathbf{e}_i'$$

$$A^{-1/2} = \mathbf{P}\Delta^{-1/2}\mathbf{P}' = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}}\mathbf{e}_i\mathbf{e}_i'$$

Properties:

1. $(\mathbf{A}^{1/2})' = \mathbf{A}^{1/2}$ (symmetric)
2. $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$
3. $\mathbf{A}^{1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1/2}\mathbf{A}^{1/2} = \mathbf{I}$

Example

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N_2 \left(\begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \right)$$

Find the following:

1. $tr(\Sigma)$
2. $det(\Sigma)$
3. eigenvalue and eigenvector of Σ
4. square root of $A = A^{1/2}$

Example: Question 1

```r
(Sigma <- cbind(c(2,1,0), c(1,2,-1), c(0,-1,2)))
```

```
#      [,1] [,2] [,3]
# [1,]    2    1    0
# [2,]    1    2   -1
# [3,]    0   -1    2
```

```r
library(matrixcalc) # load matrixcalc
matrix.trace(Sigma) # 1) compute the trace
```

```
# [1] 6
```

```r
det(Sigma) # 2) compute determinant
```

```
# [1] 4
```

```r
(r <- eigen(Sigma)) # 3) Compute the eigenvalues/eigenvect
```

```
# eigen() decomposition
# $values
# [1] 3.4142136 2.0000000 0.5857864
#
# $vectors
#            [,1]         [,2]       [,3]
# [1,]  0.5000000 7.071068e-01 -0.5000000
# [2,]  0.7071068 1.099065e-15  0.7071068
# [3,] -0.5000000 7.071068e-01  0.5000000
```

```r
l1 <- sqrt(r$values[1])
l2 <- sqrt(r$values[2])
l3 <- sqrt(r$values[3])
v1 <- r$vector[,1]
v2 <- r$vector[,2]
v3 <- r$vector[,3]
```

Compute square root of Sigma

```
# 4) Square root of A using previous formula
(Sigma.root <- l1*v1%*%t(v1) +
              l2*v2%*%t(v2) + l3*v3%*%t(v3))
```

```
#           [,1]       [,2]       [,3]
# [1,] 1.3603883  0.3826834  0.0538253
# [2,] 0.3826834  1.3065630 -0.3826834
# [3,] 0.0538253 -0.3826834  1.3603883
```

```
zapsmall(Sigma.root %*% Sigma.root)
```

```
#      [,1] [,2] [,3]
# [1,]    2    1    0
# [2,]    1    2   -1
# [3,]    0   -1    2
```

Quadratic Forms

1. $(\mathbf{X}-\mu)'\Sigma^{-1}(\mathbf{X}-\mu)$ is distributed as χ^2_p, where χ^2_p denotes the chi-square distribution with p degrees of freedom.

2. The $N_p(\mu, \Sigma)$ distribution assigns probabilty $1-\alpha$ to the solid ellipsoid
$$\{\mathbf{x} : (\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu) \leq \chi^2_p(\alpha)\}$$
where $\chi^2_p(\alpha)$ denotes the upper (100α)th percentile of the χ^2_p distribution.

3. If $p=2$, find $Pr((\mathbf{X}-\mu)'\Sigma^{-1}(\mathbf{X}-\mu) \leq 4.61)$.

```r
# See Table 3 in appendix
pchisq(4.61, df=2)
```

```
# [1] 0.9002412
```

Maximum Likelihood Estimation

Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are IID $N_p(\mu, \Sigma)$, Σ^{-1} exists. Then, the likelihood function of (μ, Σ) is

$$\begin{aligned} L(\mu, \Sigma) &= \{\text{joint density of } \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n |\ \mu,\ \Sigma\} \\ &= \prod_{i=1}^{n} \left[\frac{\exp\{-\frac{1}{2}(\mathbf{X}_j - \mu)'\Sigma^{-1}(\mathbf{X}_j - \mu)}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{1}{2}}} \right] \\ &= \frac{\exp\{-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{X}_j - \mu)'\Sigma^{-1}(\mathbf{X}_j - \mu)}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{1}{2}}} \end{aligned}$$

$\hat{\mu}$ and $\hat{\Sigma}$ are the values which maximize L.

Maximum Likelihood Estimator

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample (same as IID) from a multivariate normal population with mean μ and covariance Σ.

Then

$$\hat{\mu} = \bar{\mathbf{X}} \text{ and } \hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})' = \frac{(n-1)}{n}\mathbf{S}$$

are the maximum likelihood estimators of μ and Σ, respectively. (*Refer to pages 168 to 172 for the proof*)

mvnmle for Multivariate Normal MLE in R

```r
library(mvnmle)
library(mvtnorm)
set.seed(21)
Sigma0 <- matrix(c(1, .79, .79, 1), 2) # popn covariance
X <- rmvnorm(20, mean = c(0, 0), sigma = Sigma0) # n= 20
par.est <- mlest(X) # use mlest() to get mean/cov est
par.est$muhat
```

```
# [1] 0.1574176 0.1335548
```

```r
par.est$sigmahat
```

```
#          [,1]     [,2]
# [1,] 1.572559 1.133193
# [2,] 1.133193 1.064320
```

MLE are consistent estimators

As n increases the estimates converge to the value that the estimator is designed to estimate.

```
set.seed(21)
X <- rmvnorm(20, c(0, 0), Sigma0) # n = 20
par.est <- mlest(X)
par.est$muhat
```

```
# [1] 0.1574176 0.1335548
```

```
par.est$sigmahat
```

```
#          [,1]     [,2]
# [1,] 1.572559 1.133193
# [2,] 1.133193 1.064320
```

MLE are consistent estimators

As n increases the estimates converge to the value that the estimator is designed to estimate.

```
set.seed(21)
X <- rmvnorm(100, c(0, 0), Sigma0) # n = 100
par.est <- mlest(X)
par.est$muhat
```

```
# [1] -0.02177700 -0.01724776
```

```
par.est$sigmahat
```

```
#           [,1]      [,2]
# [1,] 0.9915466 0.8739523
# [2,] 0.8739523 1.1078480
```

MLE are consistent estimators

As n increases the estimates converge to the value that the estimator is designed to estimate.

```
set.seed(21)
X <- rmvnorm(1000, c(0, 0), Sigma0) # n = 1000
par.est <- mlest(X)
par.est$muhat
```

```
# [1] 0.07130167 0.05703656
```

```
par.est$sigmahat
```

```
#           [,1]      [,2]
# [1,] 1.0475618 0.8538465
# [2,] 0.8538465 1.0526822
```

Missing Values

The `apple` data (in `mvnmle`) provides the number of apples (in 100s) on 18 different apple trees. For 12 trees, the percentage of apples with worms (x 100) is also given.

```
data(apple) # load data
str(apple)
```

```
# 'data.frame': 18 obs. of  2 variables:
#  $ size : num  8 6 11 22 14 17 18 24 19 23 ...
#  $ worms: num  59 58 56 53 50 45 43 42 39 38 ...
```

```
# is.na() checks for missing values (NA) in each column,
# colSums(is.na()) counts the number of NA in each column
colSums(is.na(apple))
```

```
#  size worms
#     0     6
```

```r
apple.est <- mlest(apple) # use mlest() to get MLE's
apple.est$muhat
```

```
# [1] 14.72227 49.33325
```

```r
apple.est$sigmahat
```

```
#           [,1]      [,2]
# [1,]  89.53415 -90.69653
# [2,] -90.69653 114.69470
```

```r
# verify
colMeans(apple, na.rm = TRUE)  # remove rows with NA's
```

```
#     size    worms
# 14.72222 45.00000
```

```r
cov(apple, use = "complete")   # remove rows with NA's
```

```
#            size     worms
# size   84.00000 -85.09091
# worms -85.09091 111.09091
```

The Wishart Distribution

- The sampling distribution of the sample covariance matrix is called the Wishart distribution. Let

$$\mathbf{W} = (n-1)\mathbf{S} = \sum (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$$

- The dist'n of the random matrix $\mathbf{W}$ is the Wishart distribution, denoted by $W_p(n-1, \Sigma)$ with $df = n-1$.

$$\begin{aligned} \mathbf{W} &\sim W_p(n-1, \Sigma) \\ E(\mathbf{W}) &= (n-1)\Sigma \\ Cov(\mathbf{W}) &= 2(n-1)\Sigma \otimes \Sigma \end{aligned}$$

- The Wishart distribution is the multivariate antilog to the Chi-square distribution and it has similar uses.

*More information on Wishart Distribution's in this **link**.*

Wishart Dist in R

```
## Artificial
(S <- toeplitz((5:1)/5))
```

```
#      [,1] [,2] [,3] [,4] [,5]
# [1,]  1.0  0.8  0.6  0.4  0.2
# [2,]  0.8  1.0  0.8  0.6  0.4
# [3,]  0.6  0.8  1.0  0.8  0.6
# [4,]  0.4  0.6  0.8  1.0  0.8
# [5,]  0.2  0.4  0.6  0.8  1.0
```

```r
set.seed(11)
R <- rWishart(500, df = 5, Sigma = S) # generate 500 sampl
dim(R)  #  array 5 5 500
```

```
# [1]   5   5 500
```

```r
mR <- apply(R, 1:2, mean)  # ~= E[ Wish(S, 5) ] = 5 * S
round(mR, digits = 2) # round
```

```
#      [,1] [,2] [,3] [,4] [,5]
# [1,] 5.00 4.07 3.06 2.08 1.07
# [2,] 4.07 5.11 4.15 3.15 2.06
# [3,] 3.06 4.15 5.15 4.12 3.03
# [4,] 2.08 3.15 4.12 5.04 3.95
# [5,] 1.07 2.06 3.03 3.95 4.87
```

The Sampling Distribution of $\bar{\mathbf{X}}$ and $\mathbf{S}$

If $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim N_p(\mu, \Sigma)$, (IID),

- Sampling Dist'n of $\bar{\mathbf{X}}$

$$\bar{\mathbf{X}} \sim N_p(\mu, (1/n)\Sigma)$$

- Quadratic Forms

$$n(\bar{\mathbf{X}} - \mu)' \mathbf{\Sigma}^{-1} (\bar{\mathbf{X}} - \mu) \sim \chi^2_p$$

$$\mathbf{W} = (n-1)\mathbf{S} \sim W_p(n-1, \Sigma)$$

$\bar{\mathbf{X}}$ and $\mathbf{S}$ are independent.

Large-Sample Behavior of $\bar{\mathbf{X}}$ and $\mathbf{S}$

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be indep obs from a pop'n with mean μ and finite (nonsingular) covariance Σ. Then

- (Central Limit Theorem)

$$\sqrt{n}(\bar{\mathbf{X}} - \mu) \quad \text{is approx} \quad N_p(0, \Sigma)$$

- and

$$n(\bar{\mathbf{X}} - \mu)'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mu) \quad \text{is approx} \quad \chi^2_p$$

- equivalently

$$n(\bar{\mathbf{X}} - \mu)'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \mu) \quad \text{is approx} \quad \frac{p(n-1)}{n-p} F_{p,n-p}$$

for $n - p$ large.